

CLAIMS

WHAT IS CLAIMED IS:

1. A method for using extensible markup language to normalize documents, the method comprising the steps of:

determining a type of object repository storing at least one object, the object comprising metadata;

identifying the at least one object stored in the at least one object repository;

extracting at least one portion of the at least one object, wherein the at least one portion is extracted in extensible markup language (XML) format; and

transmitting the at least one portion to a processor; and

processing the at least one portion.

2. The method of claim 1, wherein some of the metadata is preserved.

3. The method of claim 2, wherein the metadata that is preserved includes at least one of author, title, subject, date created, date modified, list of modifiers, and link list information.

4. The method of claim 1, further comprising the step of:

mapping at least one field in the at least one object with a field designation identifier.

5. The method of claim 1, wherein the processor comprises at least one of a full-text engine, a metrics engine, and a taxonomy engine.

6. A system for using extensible markup language to normalize documents, the system comprising:

a determining module that determines a type of object repository storing at least one object, the object comprising metadata;

an identifying module that identifies the at least one object stored in the at least one object repository;

an extracting module that extracts at least one portion of the at least one object, wherein the at least one portion is extracted in extensible markup language (XML) format; and

a transmitting module that transmits the at least one portion to a processor; and

a processing module that processes the at least one portion.

7. The system of claim 6, wherein some of the metadata is preserved.

8. The system of claim 7, wherein the metadata that is preserved includes at least one of author, title, subject, date created, date modified, list of modifiers, and link list information.

9. The system of claim 6, further comprising:

a mapping module that maps at least one field in the at least one object with a field designation identifier.

10. The system of claim 1, wherein the processing module comprises at least one of a full-text engine, a metrics engine, and a taxonomy engine.

11. A system for using extensible markup language to normalize documents, the system comprising:

determining means for determining a type of object repository storing at least one object, the object comprising metadata;

identifying means for identifying the at least one object stored in the at least one object repository;

extracting means for extracting at least one portion of the at least one object, wherein the at least one portion is extracted in extensible markup language (XML) format; and

transmitting means for transmitting the at least one portion to a processor; and

processing means for processing the at least one portion.

12. The system of claim 11, wherein some of the metadata is preserved.

13. The system of claim 12, wherein the metadata that is preserved includes at least one of author, title, subject, date created, date modified, list of modifiers, and link list information.

14. The system of claim 11, further comprising:

mapping means for mapping at least one field in the at least one object with a field designation identifier.

15. The system of claim 11, wherein the processing means comprises at least one of a means for full-text indexing the at least one object, means for extracting metrics information from the at least one object, and means for categorizing the at least one object.

16. A processor readable medium comprising processor readable code for causing a processor to use extensible markup language to normalize documents, the medium comprising:

determining code that causes a processor to determine a type of object repository storing at least one object, the object comprising metadata;

identifying code that causes a processor to identify the at least one object stored in the at least one object repository;

extracting code that causes a processor to extract at least one portion of the at least one object, wherein the at least one portion is extracted in extensible markup language (XML) format;

transmitting code that causes a processor to transmit the at least one portion to a processor; and

processing code that causes a processor to process the at least one portion.

17. The medium of claim 16, wherein some of the metadata is preserved.

18. The medium of claim 17, wherein the metadata that is preserved includes at least one of author, title, subject, date created, date modified, list of modifiers, and link list information.

19. The medium of claim 16, further comprising:

mapping code that causes a processor to map at least one field in the at least one object with a field designation identifier.

20. The medium of claim 16, wherein the processing code comprises at least one of a full-text engine, a metrics engine, and a taxonomy engine.